



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2018

Comparing cultural differences with domain-specific differences of appreciating and understanding values

Christen, Markus

Abstract: Human endeavors such as promoting science or creating art have a close connection to underlying values and virtues. For example, creativity, curiosity or objectivity provide orientation when working as scientist or artist, outlining the ultimate aim of practitioners active in those domains. The appreciation and interpretation of domain values are expected to depend upon the domain, which should less be the case for moral values. This study investigates differences in the semantics, perceived importance and interpretation of 10 domain and 10 moral values in two domains (art and medicine) and two cultural settings (US, N = 336; Switzerland, N = 554). It is shown that the semantic understanding of values is robust with respect to culture and domain and that the appreciation of values varies in dependence of the domains but not the culture. Cultural factors have a greater impact compared to the domain for domain value interpretation, but not moral value interpretation.

DOI: <https://doi.org/10.1080/03057240.2018.1469477>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-160327>

Journal Article

Accepted Version

Originally published at:

Christen, Markus (2018). Comparing cultural differences with domain-specific differences of appreciating and understanding values. *Journal of Moral Education*, 47(3):333-345.

DOI: <https://doi.org/10.1080/03057240.2018.1469477>

Comparing Cultural with Domain-Specific Differences of Appreciating and Understanding Values

Markus Christen¹

¹ Institute of Biomedical Ethics and History of Medicine, University of Zurich, Switzerland

Correspondence concerning this article should be addressed to Markus Christen, Institute of Biomedical Ethics and History of Medicine, Winterthurerstrasse 30, 8006 Zurich, Switzerland, christen@ethik.uzh.ch

Acknowledgement: The author thanks Darcia Narvaes (University of Notre Dame) and Carmen Tanner (University of Zurich) for their valuable input when preparing this study and Markus Kneer for advice regarding the statistical analysis.

Abstract: Human endeavors such as promoting science or creating art have a close connection to underlying values and virtues. For example, creativity, curiosity or objectivity provide orientation when working as scientist or artist, outlining the ultimate aim of practitioners active in those domains. The appreciation and interpretation of domain values are expected to depend upon the domain, which should less be the case for moral values. This study investigates differences in the semantics, perceived importance and interpretation of 10 domain and 10 moral values in two domains (art and medicine) and two cultural settings (USA, N=336; Switzerland, N=554). It is shown that the semantic understanding of values is robust with respect to culture and domain and that the appreciation of values varies in dependence of the domains but not the culture. Cultural factors have a greater impact compared to the domain for domain but not moral value interpretation.

Keywords: Values; Virtues; Art; Medicine; Semantics; Culture; Practice

Biography: Markus Christen is a Senior Research Fellow in the Institute of Biomedical Ethics and History of Medicine at the University of Zurich and Managing Director of the “Digital Society Initiative” of the University of Zurich. His research focus is empirical ethics, neuroethics, ethics & technology, Serious Moral Games and data analysis methodologies.

Comparing Cultural with Domain-Specific Differences of Appreciating and Understanding Values

Abstract: Human endeavors such as promoting science or creating art have a close connection to underlying values and virtues. For example, creativity, curiosity or objectivity provide orientation when working as scientist or artist, outlining the ultimate aim of practitioners active in those domains. The appreciation and interpretation of domain values are expected to depend upon the domain, which should less be the case for moral values. This study investigates differences in the semantics, perceived importance and interpretation of 10 domain and 10 moral values in two domains (art and medicine) and two cultural settings (USA, N=336; Switzerland, N=554). It is shown that the semantic understanding of values is robust with respect to culture and domain and that the appreciation of values varies in dependence of the domains but not the culture. Cultural factors have a greater impact compared to the domain for domain but not moral value interpretation.

Keywords: Values; Virtues; Art; Medicine; Semantics; Culture; Practice

The empirical investigation of values has gained interest in a time where many conflicts between individuals, groups or cultures are framed as conflicts of values (Huntington, 1993; Pearce & Littlejohn, 1997; Schwartz & Sagie, 2000). Commonly understood as abstract and desirable standards (Rokeach, 1973; Schwartz, 1992), values are proposed to guide decision-making and behavior. Furthermore, they create a frame for institutionalized settings such as professions, where values provide orientation, outlining the ultimate aim of practitioners active in those domains. Which values are considered important in a practice domainⁱ likely depends upon the characteristics of the domain. Other types of values, in particular moral values, are expected to be universal in the sense that they should be upheld across cultures and domains (Hare, 1954; Taylor, 1978). However, moral relativists refuse these universalist claims based on empirical and conceptual grounds (Shweder, 1993; Quintelier & Fessler, 2012).

There is a rich empirical literature in social and evolutionary psychology that investigates the origin and interrelation of different types of values as a function of cultural change (e.g., Inglehart & Baker, 2000; Vauclair, Hanke, Fischer & Fontaine, 2011), as well as the question to what extent in particular moral values

are influenced by cultural factors (e.g., Nisan, 1987; Haidt, 2007; Krebs, 2008). The latter point obviously relates to the longstanding issue, whether morality is universal or culturally determined. Whereas moral universalists often refer to the adaptive benefit of social cooperation, thus providing the foundation that at least some moral values are culturally stable, cultural psychologists argue that judgments of right and wrong are influenced by culturally based conceptions of the self (e.g. Shweder, Much, Mahapatra, & Park, 1997). The current state of knowledge in social and evolutionary psychology points to a rather complex picture with respect to differences in understanding (moral) values. For example, Vauclair and Fischer (2011) concluded based on data of the World Value Survey that not all values that usually are understood as moral values satisfy claims of universality. They found that justice, fairness and social cooperation were judged in a nearly universal way; whereas issues of individual's duties and obligations to the social group are influenced by cultural models of the self.

However, fewer studies seem to exist that assess whether the understanding of moral values differs across practice domains within the *same* culture.ⁱⁱ This study intends to add findings whether not only culture, but also professional orientation (i.e., attachment to certain domains associated with professional standards) may influence the appreciation and understanding of moral and non-moral values. The results may thus inform the debate regarding the degree of universality of moral values. In addition, in line with the Special Issue, the aim of the study is to inform the research of the “Developing Virtues in the Practice of Science” project that investigates dispositions cultivated by scientific practice using tools of psychology and anthropology and that seeks to illuminate philosophical understanding of those dispositions framed as virtues. The scientific context in this project refers to the biomedical sciences, whereas artists serve as a comparison group. This is why the study presented here focuses on two practice domains, art and medicine.

The goal of this study is to investigate differences in the *semantic understanding* (which synonyms would one use to best circumscribe a value term?), the *appreciation* (which values are or should be considered important in which domain?) and the *interpretation* (along four interpretation dimensions referring to the moral character, emotional appeal, official status and community-orientation of the values) of moral and domain values, whereas latter likely characterize important practice orientation dimensions in either domain. We investigate those possible differences for persons with an attachment to different practice backgrounds (art and medicine) and emerging from different cultural settings (USA and Switzerland). The hypotheses of the study are as follows:

- a) The semantic understanding of both types of values differ across domain and culture. The motivation for this hypothesis is that Christen et al. (2016) demonstrated remarkable differences in the

value maps emerging from Thesaurus data with respect to the “semantic neighborhood” of value terms.

- b) The appreciation of domain and moral values should differ across domains but not strongly across cultures. The motivation for this hypothesis is that values are expected to provide a unique framing for a certain practice domain; e.g., it is very plausible that “beauty” is considered more relevant in art compared to medicine. We also expect that medicine has a stable “appreciation spectrum” of values across cultures given that the professional ethos of medicine is dominated by global frameworks such as the Hippocratic oath or the Helsinki declaration.
- c) Domain values demonstrate more variability in interpretation dimensions with respect to both culture and domain compared to moral values. The motivation of this hypothesis is that various studies point to cultural differences with respect to non-moral values, whereas moral values seem to be more stable with respect to cultural influences (see the references in the introduction as examples). Furthermore, the study of Christen et al. (2014) showed remarkable differences in the interpretation of values also when comparing practice domains (medicine vs. business in this case) within a single culture.

Originally, this study was not designed in view of the “Developing Virtues in the Practice of Science” project. Rather, the data of this work emerge from a larger follow-up study of research that created a comprehensive list of 78 value groups using thesaurus databases. In this research (Christen et al., 2016), we used a methodology that integrates empirical value research with linguistics and novel computer visualization tools to map and visualize value spaces. Based on 460 value terms both for US-English and German, we created for each language a map of 78 value classes. Thus, there is no complete matching with respect to the practice domains and values used in this study compared to the virtue project. A subset of 20 out of those 78 values have been used for this study, and the practice domain is medicine rather than biomedical research.

The study that informed this research was based on thesaurus databases that help its users find words related to a concept but having slightly different shades of meaning or connotation. As outlined in Christen et al. (2016), the thesaurus data allowed creating value maps that show the “semantic neighborhood” of values; i.e., provide a measure how “similar” two values are in terms of synonyms used to circumscribe it. In this previous study, we found remarkable differences when comparing semantic neighborhoods of values in English and German, pointing to cultural differences in language use. As Thesaurus dictionaries have a long tradition, starting in the 17th century and cumulated in famous books like Roget’s “Thesaurus”, published in 1852 (Hüllen, 2004), they reflect language use over a considerable long period across a large social spectrum. However, the participants involved in this study emerge from a less comprehensive background, in

particular regarding the Swiss sample. Findings regarding cultural differences thus cannot be extrapolated to the overall German-speaking language area.

The structure of the paper is as follows. The next section outlines the methodology used in detail (referring to the larger project that contained a survey study of 78 values). Then the results of the survey are presented. In the final section, the relevance of the findings for the “Developing Virtues in the Practice of Science” project are discussed.

Method

The data used in this study emerge from a larger, ongoing research project that is a follow-up of Christen et al. (2016). The sections “Procedure” and “Participants” describe the methodology used in the full survey, whereas the section “Measures” refers to this study.

Procedures

Starting position of the research is the list of 78 values and value synonym terms generated in Christen et al. (2016). Those value synonyms sets were composed of in total 440 (English) and 449 (German) terms, whereas for each term we have a quantitative rating describing the affinity of the term to its set. Using a subset of those terms, we created an online survey that consisted of the following parts: First, the participants provided some general demographic information (gender, age, country in which the participant is born and in which the participant is currently living). Then, the participants attributed themselves to one out of four domains to which the participants had the highest practice affinity (art, business, medicine, education; the participants were provided with detailed information on how to assess their degree of affinity). We also rated the experience of the participants in each domain using six different qualifiers (training, internship experience, work experience, passive interest like domain-specific reading, concrete domain-specific activities, number of hours per week used for domain-related activities; the qualifiers were similar in each domain). It is important to mention that this attribution is rather loose, i.e. is based on the best fit given a predetermined set of practice domains. Not all persons who attributed themselves to “medicine” are medical professionals and not all persons who attributed themselves to “art” can be expected to have a similar understanding of what “art” means. Thus, the persons are rather “practitioners” than “professionals”, although their degree of attachment to the respective domain is captured by an experience index (see below and endnote i) that also includes information, whether, e.g., the persons have a formal training in the domain.

The domain descriptions in the survey were broad. For example, in the case of art, the description was as follows: “Training as artist, dancer, musician, writer and the like; work experience in cultural institutions of

any kind (theater, museum etc.); own experience in creating pieces of art (any kind, including music); interest in art and culture (e.g., being an art collector).” Thus, one cannot assume that the persons have a similar understanding of the domain.

After that, the participants were asked to “construct” their value based on the procedure described below; a random sample of 12 (USA) respectively 10 (CH) out of 78 value synonyms sets has been presented per participant. The “self-constructed value” was then assessed by the participant along four different dimensions (see below) and evaluated with respect to its importance of the chosen domain using a 6-point Likert scale. The “*self-construction*” of a value worked as follows: The participants were presented with 7-14 terms per set. Then they were asked to decide which terms belong to the same value category, and (by using the mouse pointer) to drag all terms that do not fit to their understanding of the value to a “trash bin”. The remaining terms then represent the value based on their choices. All participants performed a simple example under instructions to ensure that they understood the procedure.

For example, the participants were presented with the following list of terms (in an alphabetical order): autarchy, autonomy, independence, purity, self-determination, self-government, self-reliance, self-rule. By using their mouse, they were asked to drag all terms that did not fit into the box “trash”, such that the remaining terms form, according to their opinion, a synonym set circumscribing a value. One term in each set (“purity” in this case) was a distractor term, i.e., a term that was not a synonym to the other set members and that was expected to be dragged to the trash bin. Distractor terms emerged from the value map described in Christen et al., 2016 such that the distractor terms had a large semantic distance from the synonymy set under consideration; they were used for the data quality check (see below). For naming the value synonyms sets, we used the names as described in Christen et al. (2016) – the example above refers to the value “autonomy”. Checks performed showed that the terms used as names were usually also those terms people deleted only rarely.

The *interpretation dimensions* referred to different facets on how the values that has been self-constructed by the participants following the method described above can be appreciated: The first dimension refers to the “moral – non-moral” distinction; i.e. participants had to evaluate to what extent the self-constructed value claims to be universally valid and whether actions corresponding to the self-constructed value are judged as right or wrong. The second dimension refers to the degree of emotionality the self-constructed value evokes, i.e. whether following this value feels (completely) right or (completely) wrong for the participant. The third dimension assesses the official character of the self-constructed value, i.e. whether the value relates to the compliance to given internal rules or laws or not in the respective domain to which the

person has attributed herself. The fourth dimension evaluates the community-orientation of the self-constructed value, i.e. the participants are asked to what degree the value supports the cooperative being-together of people. The Likert scale range for the four interpretation dimensions was between 1 and 6; the lower the number, the more moral etc. is the value.

The *importance* of the value of the domain to which the participant has attributed herself was assessed in two questions: how important the value is and how important the value should be in the domain. Again the Likert scale range was between 1 and 6; the higher the number, the higher is the perceived or desired importance of the value.

Participants

A US and a Swiss sample of participants was approached in two waves each. The survey among participants for the English value synonyms sets was conducted using Amazon Mechanical Turk (restricted to US Americans). In this sampling method, participants received a small reimbursement for completion of the survey. The survey for the German value synonyms sets was conducted using a service provided by the University of Zurich that addresses students, faculty and staff of the university (i.e., Swiss-German speaking). Here, a lottery (book vouchers) was chosen as incentive for participation. The information about origin and current place of residence was used to ensure that the participant's first language is English or German. Using this information does not exclude the possibility that persons whose first language is not English or German still could have participated in the survey; however, the distribution channels (Amazon Mechanical Turk, restricted to US citizens, respectively persons born in Switzerland and associated to a German-speaking university) makes it reasonable to assume that only very few participants have a foreign language as first language.

We then used a between-group-comparison-design of four groups: US participants that attributed themselves to art (US-ART) or medicine (US-MED), and Swiss participants that attributed themselves to art (CH-ART) or medicine (CH-MED). We compared either between cultures (same practice domain) or between practice domains (same culture); we did not compare between different cultural and practice backgrounds simultaneously (i.e., no comparisons US-ART vs. CH-MED or US-MED vs. CH-ART).

The study was cleared in accordance with the ethical review processes of the University of Zurich (see <https://www.phil.uzh.ch/de/forschung/ethik.html> for details) and following the “Ethical Guidelines for Psychologists of the Swiss Society for Psychology”.

Measures and Statistics

A first measure refers to *quality control*. The data reveal that not all distractor terms used served their purpose. In some cases the term set included terms that are very rarely used in everyday language – and people tended to delete uncommon terms. This was the case in eight English and four German value synonyms sets – those sets were not used in the quality check. Performing the quality check then led to the exclusion of 20.7% of all English participants (724 out of 913 remained) and of 15.5% of all German participants (1056 out of 1250 remained). Of the remaining participants, 336 US and 554 Swiss participants attributed themselves to the domains Art (US N=221, CH N=290) or Medicine (US N=115, CH N=264). Those formed the sample used in this study. Per group, we had in the mean between 18-34 single ratings per value for the US sample and 34-37 single ratings for the Swiss sample.

The *experience index* was created as follows: Each of the first five qualifiers (training, internship experience, work experience, passive interest like domain-specific reading, concrete activities) ticked yielded one point. For the number of hours per week used for domain-related activities, the possible choices and associated points were as follows: less than 5 hours (1), 5 to 10 hours (2), 11 to 20 hours (3), 21 to 40 hours (4), more than 40 hours (5). All points were added, such that the range of the experience index per person was between 1 and 10. To check the sensitivity of the result with respect to calculating the experience index, the weight of the time used for domain related activities was also multiplied with a factor of either 0.5 or 2. In the result section, influences of this sensitivity analysis are discussed.

The analysis was restricted to 20 out of 78 *values*. The focus was on values that are either potentially relevant for at least one of the two practice domains (those values were autonomyⁱⁱⁱ, beauty, carefulness, creativity, intelligence, objectivity, openness, precision, progress, wisdom) or moral values (those values were altruism, care, cooperation, dignity, fairness, honesty, integrity, respect, responsibility, solidarity). The choice was informed by the values that are studied in the “Developing Virtues in the Practice of Science” project, as all values in this study are explicitly or implicitly^{iv} investigated in latter project. Furthermore, the morality-dimension assessed in the survey provides a justification for calling the second group “moral values”, as they received significant higher values in the “moral – non-moral” dimension compared to the domain values (overall mean moral values: 1.98; overall mean domain values: 3.51, $p < .001$; Mann-Whitney-Test). It is reminded that the set of possible values was restricted by the original set of 78 values, as the data originally were collected to evaluate further the value map identified in Christen et al. (2016) and not to test the appreciation of all relevant values for the domains chosen. This means that the chosen set is not optimal as there are likely other values that may be relevant for either domain but that are not covered by this study. The supplementary information contains all synonyms for those 20 values for both languages.

The *semantic understanding* of a value was determined by computing the frequency of conservation of terms (excluding the distractor) of a synonyms set, i.e. how often a term of the set was considered to be part of the understanding of the value. This was done for each of the four groups (US-ART, US-MED, CH-ART, CH-MED) separately. In this way, for each value and group, we obtained a sequence of numbers representing the frequency of terms in the synonym set between 0 and 1, whereas 0 indicates that the term was deleted in all cases and 1 indicates that the term was always kept when circumscribing the value. To illustrate this for the autonomy example given above, the conservation frequencies (provided here in percentages) of the terms were: autarchy: 58.3%, autonomy: 94.4%, independence: 97.2%, self-determination: 97.2%, self-government: 91.7%, self-reliance: 97.2%, self-rule: 94.4%. We used Pearsons Chi-Square test (Yates correction, as the number of ratings per value was below 50) for making statistical comparisons between the conservation distributions of the same values of two different groups.

Group comparisons with respect to *interpretation* (2 dimensions) and *appreciation* (4 dimensions) of the values were calculated using the nonparametric Mann Whitney test, because most of the dimension data are not normally distributed. For calculating the group comparisons, the statistical problem of multiple comparisons has to be taken into account. The number of possible comparisons is 4 (groups) x 20 (values) x 6 (dimensions) = 480; i.e. there is a considerable likelihood to commit a type 1 error (i.e.; some of the identified significant pairwise differences will appear by chance). However, standard corrections for multiple comparison such as the Bonferroni or Holm correction (Holm 1979) tend to be very conservative, in particular if the number of pairwise comparisons is large such as in this case – meaning that the likelihood to commit a type 2 error (i.e.; missing significant differences) is high (Howell, 2013). However, in our case, we are not interested whether *certain* pairs yield statistical significant differences, but whether the *overall number* of statistical significant differences is different when comparing cultures vs. domains.^v Because both the occurrence of type 1 and type 2 errors is equally likely in either case, the null hypothesis that the probability of finding statistical different pairwise comparisons is equal in all domains remains valid. Statistical significant differences when comparing the number of significant pair-differences across domains and cultures can be assessed using Pearsons Chi-Square test. For better balancing the likelihoods of type 1 and type 2 errors, the suggestion of Benjamin et al. (2017) was used; i.e., the threshold for significance in the individual pairwise comparison was set at $p < 0.005$. All calculations were performed using Mathematica® Version 11.1.

Results

Because group comparisons are the primary target of this study, demographic differences between the four groups are reported first (Table 1). We find that females (except US-ART) form a majority in the groups, the gender distribution differences are significant when comparing across languages and (only for the US data) across practice domains. We also find significant age differences between languages, but not for practice domains. With respect to the experience index, the group CH-MED is more experienced compared to CH-ART and US-MED; the result is robust with respect to the sensitivity analysis for the experience index. This means that group differences found in the other measures could in some cases also be influenced by gender, age or experience effects. This will be taken into account when discussing the results.

Table 1. Study sample description

| | US sample (N=336) | | Swiss sample (N=554) | |
|------------|-------------------|----------------|----------------------|----------------|
| | US-ART (N=221) | US-MED (N=115) | CH-ART (N=290) | CH-MED (N=264) |
| % Female | 43.9% | 56.5% | 68.6% | 71.6% |
| Mean Age | 30.4 | 33.1 | 28.8 | 31.0 |
| Experience | 4.74 | 5.08 | 4.72 | 5.98 |

In the group comparisons for the *appreciation* of values, hypothesis b) predicts that the number of significant pair comparisons is higher when comparing across domains instead when comparing across cultures. Indeed, there are 33 significant pair-comparisons across domains, but only 5 significant pair-comparisons across cultures (Table 2); this is a significant deviation from the null hypothesis ($p < 0.001$). There are no significant differences when comparing moral and domain values. This is compatible with hypothesis b).

Table 2. Number of statistical significant differences with respect to the appreciation dimensions (is and ought; $p < 0.005$).

| | # of significant differences | | |
|----------------------------|------------------------------|----|-----------------|
| Comparisons across culture | 5 | 2 | Moral values |
| | | 3 | Practice values |
| Comparisons across domains | 33 | 17 | Moral values |
| | | 16 | Practice values |

As an illustration, Figure 1 outlines the results in the cross-culture comparison for the “is”-dimension, where the number of significant pairwise differences was the lowest (only “altruism” for art and “objectivity” for medicine were different across cultures in the same domain). When comparing the absolute ratings across

domains, it is remarkable that in the medical domain, the overall scores of the values are higher compared to art and that the relevance of moral values (printed in red) is much stronger emphasized. In medicine, only one value (“objectivity”) yielded significant differences in both appreciation dimensions, which is compatible with the assumption of a stable “appreciation spectrum” of values in medicine (as indicated in hypothesis b).

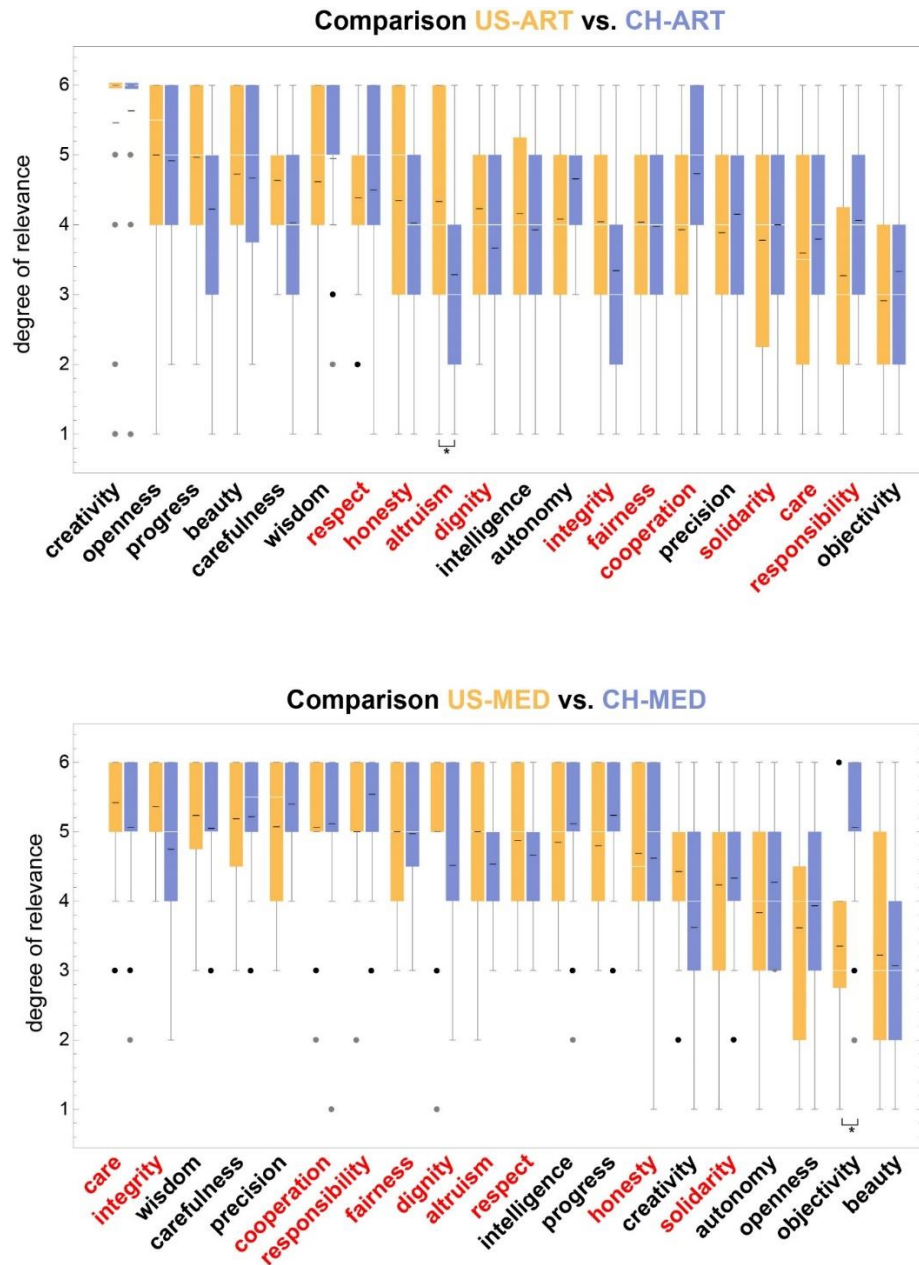


Figure 1. Box-plots that compare the degree of relevance (how important is the value) in the domains art (above) and medicine (below). Red typeset moral values. *: $p < 0.005$.

In the group comparison for the interpretation dimensions, hypothesis c) predicts that the number of significant pair comparisons is higher for domain values both when comparing across cultures and when comparing across domains. Indeed, we find 13 significant pair-comparisons across cultures out of which 10 are attributed to domain values, but only 2 significant pair-comparisons across cultures (Table 3). For the cultural comparison, this is a significant deviation from the null hypothesis ($p < 0.05$), but not for the domain comparison. This is partially compatible with hypothesis c).

Table 3. Number of statistical significant differences with respect to the four interpretation dimensions (moral, emotional, rule-oriented, community-oriented; $p < 0.005$).

| | # of significant differences | | |
|----------------------------|------------------------------|----|-----------------|
| Comparisons across culture | 13 | 3 | Moral values |
| | | 10 | Practice values |
| Comparisons across domains | 2 | 1 | Moral values |
| | | 1 | Practice values |

Discussion

The study investigated how cultural and practice background influences the semantic understanding, appreciation and interpretation of values. We find that hypothesis a) is not confirmed, that hypothesis b) is confirmed and that hypothesis c) is partially confirmed, which yields the following summary: The overall semantic understanding of values, i.e. their meaning expressed by determining which synonyms are conserved when circumscribing a value, seems to be quite robust with respect to practice background and culture. This can be seen as an indication that the people roughly have a common understanding, what values like “creativity”, “honesty” or “precision” mean. When people then assess the appreciation of values in different domains, we indeed find only very few cultural differences across the same domain, but many differences within the same culture and different domains. However, with respect to the interpretation of values, we do see remarkable differences, in particular across cultures. However, this mainly affects domain values, but not moral values. This finding is compatible to the moral universalist position as outlined in the introduction.

It is of particular interest that hypothesis a) has not been confirmed given that the study of Christen et al. (2016) yielded remarkable differences with respect to the semantic neighborhood of value terms. Whether this undermines the results of the previous study is, however, not clear. The result could also indicate that the task used in this study – deleting synonyms that do not fit based on the person’s understanding – might

not be an adequate measure for differences seen in the value map. Alternatively, the measure used for interpreting the data – determining the frequency of conservation of synonymy terms – may not be sufficiently capturing semantic differences. One could also understand each synonym set as a network of terms and calculate the pairwise co-occurrence of terms in each set, leading to a distribution of numbers representing the edge weight in the network. A statistical analysis of those distributions might be a better indicator of semantic differences. However, analyzing this aspect goes beyond the scope of this paper and is subject of further research.

Another important point is that the groups differ to some degree with respect to demographic characteristics. The age differences are, although significant, rather small and likely not relevant for influencing the results. However, the gender differences are remarkable and require a deeper investigation. Therefore, for each measure, it was calculated whether significant differences between male and female participants are found in all four groups (US-ART, US-MED, CH-ART, CH-MED) for each value separately using the stronger criterion for statistical significance as outlined in the methods section. No indication was found that gender is a relevant factor both in the appreciation and the interpretation dimension.

As the CH-MED sample is significantly more experiences than the other three groups using our experience index, the correlation of all measures with experiences were also calculated for each single value (using Spearman rank correlation) and we checked whether significant correlations are found (again using $p < 0.005$ as significance criterion and taking the sensitivity analysis into account. It was found that differences in experiences are not likely to influence the group differences identified in this study. Only in one value (“intelligence”) we found robust indications that experience is influencing the appreciation dimension (correlation US-ART: 0.51, correlation CH-ART: 0.51; $p < 0.005$, robust result when performing the sensitivity analysis).

It is important to point to the limitations of this study. First, the two sampling procedures are different and may yield differences with respect to the personal background of the persons not captured by the few demographic questions asked. This is because Amazon Mechanical Turk is known to have a higher demographic diversity compared to university student populations (Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010); and although the Swiss sampling procedure goes beyond students, it is still restricted to a rather academic setting, as it approaches persons with a professional affiliation to the University of Zurich. Second, although the overall number of approached persons is high, the ratings per value in the four groups is small for the group US-MED, which certainly limits the validity of the results. Third, in order to obtain sufficient numbers per domain, the number of domains was restricted and persons were forced to attribute themselves to one of the four domains to which they have the closest affinity, although their real professional experience

is in yet another domain. Some domains such as medicine are likely to attract persons who indeed have a professional relation to medicine, whereas the domain art likely includes also persons with another professional background not present in the selection, but still have a relation to art as practitioner. This limits the significance of the domain comparisons. Fourth, the intercultural comparison is likely influenced by ambiguity in translation (the “indeterminacy of translation thesis”; Quine, 1960, chap. 2), i.e. there is no guarantee that the terms used in the synonyms sets of a certain value are perceived as one-to-one translations by the participants. Finally, we remind that the two cultural settings chosen are still rather close (i.e., both belong to the Western context) and that the Swiss sample does not allow extrapolating the results to the entire German-speaking area. Comparisons involving two cultures that are more distant are likely to generate different results.

With respect to the relevance of those results for the “Developing Virtues in the Practice of Science” project, the following point can be emphasized: Science – or at least the ethos of science – is generally understood to be a global endeavor dominated by virtues that should transcend cultural borders. The same can be said for medicine. However, our results indicate that cultural aspects do to some degree influence the interpretation and, to a lesser extent, the appreciation of values. Survey studies conducted in the project in order to investigate the virtues cultivated by scientific practice – that is, the cognitive, behavioral, and emotional dispositions to act in ways that advance the good of both the individual and a given community – will have to consider such differences. The empirical work in this project will have to find ways to evaluate whether the persons interrogated do actually attach the same moral and emotional flavor to values such as creativity, integrity or progress that are undisputedly important orientations for cultivating virtue in science.

References

- Benjamin, D.J., Berger, J.O., Johannesson M., et al. (2017). Redefine statistical significance. *Nature Human Behavior*. Published online, 01. September 2017, DOI: 10.1038/s41562-017-0189-z
- Christen, M., Ineichen, C., Tanner, C. (2014). How moral are the principles of biomedical ethics? *BMC Medical Ethics*, 15: 47.
- Christen, M., Narvaez, D., Tanner, C., Ott, T. (2016). Mapping Values: Using Thesauruses to Reveal Semantic Structures of Cultural Moral Differences. *Cognitive Systems Research*, 40, 59-74.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998-1001.

Hare, R. M. (1954). Universalisability. *Proceedings of the Aristotelian Society*, 55, 295–312.

Howell, D.C. (2013). *Statistical methods for psychology*. Belmont, Calif.: Wadsworth

Hüllen, W. (2004). *A history of Roget's thesaurus*. Oxford: Oxford University Press.

Huntington, S. P. (1993). The clash of civilizations? *Foreign Affairs*, 72(3), 22–49.

Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*. Special Issue: Looking forward, looking back: Continuity and change at the turn of the millenium, 65, 19–51.

Krebs, D. L. (2008). Morality: An evolutionary account. *Perspectives on Psychological Science*, 3, 149–172.

MacIntyre, A. (1985). *After virtue: A study in moral theory* (2nd ed.). London: Duckworth

Nisan, M. (1987). Moral norms and social conventions: A cross-cultural comparison. *Developmental Psychology*, 23, 719–725.

Pearce, B., & Littlejohn, S. W. (1997). *Moral conflict: When social worlds collide*. Thousand Oaks, CA: Sage Inc.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Quintelier, K., & Fessler, D. T. (2012). Varying versions of moral relativism: The philosophy and psychology of normative relativism. *Biology and Philosophy*, 27(1), 95–113.

Rokeach, M. (1973). *The nature of human values*. New York: The Free Press.

Ross, J., Irani, I., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers? Shifting demographics in Amazon Mechanical Turk. In *Proceedings of the conference on human factors in computing systems*, Atlanta, GA, April 10–15 (pp. 2863–2872).

Schwartz, S. H., & Sagie, G. (2000). Value consensus and importance: A cross-national study. *Journal of Cross-Cultural Psychology*, 31(4), 465–497.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). New York: Academic Press.

Shweder, R. A. (1993). The cultural psychology of emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 417–431). New York: Guilford.

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The “Big three” of morality (autonomy, community, and divinity), and the “Big three” explanations of suffering. In A. Brandt, & P. Rozin (Eds.), *Morality and health* (pp. 119–169). New York: Routledge.

Taylor, P. W. (1978). On taking the moral point of view. *Midwest Studies in Philosophy*, 3(1), 35–61.

Vauclair, C.-M., Hanke, K., Fischer, R., & Fontaine, J. R. J. (2011). The structure of human values at the culture level: A meta-analytical replication of Schwartz’s value orientations using the Rokeach Value Survey. *Journal of Cross-Cultural Psychology*, 42, 186–205.

Vauclair, C.-M., & Fischer, R. (2011). Do cultural values predict individuals’ moral attitudes? A cross-cultural multilevel approach. *European Journal of Social Psychology*, 41, 645–657.

ⁱ In this contribution, the notion of “practice domain” refers to coherent and complex forms of socially established cooperative human activity through which goods internal to that form of activity are realized following certain standards of excellence (MacIntyre, 1985), whereas these standards can be understood as values. Those domains are not necessarily related to clearly defined professional curricula and the like. For persons active in those domains, the notion of “practitioner” is used.

ⁱⁱ An exploratory search in the database PsychINFO using the search expression [“moral value” AND difference AND (social OR professional)] yielded no relevant paper. If the search is not restricted to moral values, the number of entries increases largely, yet many papers use the term “value” in a different meaning.

ⁱⁱⁱ Autonomy is a special case: although the values is considered to be an important ethical orientation in medicine, empirical studies show a significant lower appreciation of autonomy as moral value (see Christen et al. 2014 for details) – this was also the case in this sample. We therefore grouped autonomy into the practice domain.

^{iv} For example, the “Developing Virtues in the Practice of Science” project investigated the value “meticulousness” defined as “attention to method and details in one’s conduct”, which can reasonably be related to the notion of “carefulness” used in this study.

^v It can be added that the clear difference when comparing domains vs. cultures (results shown in Tables 2 and 3) remains even when taking the Bonferroni correction – arguably, the most conservative one and the one with the highest risk of Type 2 errors – although the numbers of significant pairs themselves change.